Better Explained Friendly lessons, lasting insight

Search...

An Intuitive (and Short) Explanation of Bayes' Theorem

Bayes' theorem was the subject of a detailed article. The essay is good, but over 15,000 words long — here's the condensed version for Bayesian newcomers like myself:

- **Tests are not the event.** We have a cancer *test*, separate from the event of actually having cancer. We have a *test* for spam, separate from the event of actually having a spam message.
- **Tests are flawed.** Tests detect things that don't exist (false positive), and miss things that do exist (false negative).
- **Tests give us test probabilities, not the real probabilities.** People often consider the test results directly, without considering the errors in the tests.
- **False positives skew results.** Suppose you are searching for something really rare (1 in a million). Even with a good test, it's likely that a positive result is really a *false positive* on somebody in the 999,999.
- **People prefer natural numbers.** Saying "100 in 10,000″ rather than "1%" helps people work through the numbers with fewer errors, especially with multiple percentages ("Of those 100, 80 will test positive" rather than "80% of the 1% will test positive").
- **Even science is a test**. At a philosophical level, scientific experiments can be considered "potentially flawed tests" and need to be treated accordingly. There is a *test* for a chemical, or a phenomenon, and there is the *event* of the phenomenon itself. Our tests and measuring equipment have some inherent rate of error.

Bayes' theorem finds the actual probability of an event from the results of your tests. For example, you can:

- **Correct for measurement errors**. If you know the real probabilities and the chance of a false positive and false negative, you can correct for measurement errors.
- **Relate the actual probability to the measured test probability.** Bayes' theorem lets you relate Pr(A|X), the chance that an event A happened given the indicator X, and Pr(X|A), the chance the indicator X happened given that event A occurred. Given mammogram test results and known error rates, you can predict the actual chance of having cancer.

Anatomy of a Test

The article describes a cancer testing scenario:

- 1% of women have breast cancer (and therefore 99% do not).
- 80% of mammograms detect breast cancer when it is there (and therefore 20% miss it).
- 9.6% of mammograms detect breast cancer when it's **not** there (and therefore 90.4% correctly return a negative result).

Put in a table, the probabilities look like this:

	Cancer (1%)	No Cancer (99%)
Test Pos	80%	9.6%
Test Neg	20%	90.4%

How do we read it?

- 1% of people have cancer
- If you **already have cancer**, you are in the first column. There's an 80% chance you will test positive. There's a 20% chance you will test negative.
- If you **don't have cancer**, you are in the second column. There's a 9.6% chance you will test positive, and a 90.4% chance you will test negative.

How Accurate Is The Test?

Now suppose you get a positive test result. What are the chances you have cancer? 80%? 99%? 1%?

Here's how I think about it:

- Ok, we got a positive result. It means we're somewhere in the top row of our table. Let's not assume anything — it could be a true positive or a false positive.
- The chances of a *true positive* = chance you have cancer * chance test caught it = 1% * 80% = .008
- The chances of a *false positive* = chance you don't have cancer * chance test caught it anyway = 99% * 9.6% = 0.09504

The table looks like this:

	Cancer (1%)	No Cancer (99%)
Test Pos	True Pos: 1% * 80%	False Pos: 99% * 9.6%
Test Neg	False Neg: 1% * 20%	True Neg: 99% * 90.4%

And what was the question again? Oh yes: what's the chance we really have cancer if we get a positive result. The chance of an event is the number of ways it could happen given all possible outcomes:

```
Probability = desired event / all possibilities
```

The chance of getting a real, positive result is .008. The chance of getting any type of positive result is the chance of a true positive plus the chance of a false positive (.008 + 0.09504 = .10304).

So, our chance of cancer is .008/.10304 = 0.0776, or about 7.8%.

Interesting — a positive mammogram only means you have a 7.8% chance of cancer, rather than 80% (the supposed accuracy of the test). It might seem strange at first but it makes sense: the test gives a false positive 10% of the time, so there will be a **ton** of false positives in any given population. There will be so many false positives, in fact, that **most** of the positive test results will be wrong.

Let's test our intuition by drawing a conclusion from simply eyeballing the table. If you take 100 people, only 1 person will have cancer (1%), and they're nearly guaranteed to test positive (80% chance). Of the 99 remaining people, about 10% will test positive, so we'll get roughly 10 false positives. Considering all the positive tests, just 1 in 11 is correct, so there's a 1/11 chance of having cancer given a positive test. The real number is 7.8% (closer to 1/13, computed above), but we found a reasonable estimate without a calculator.

Bayes' Theorem

We can turn the process above into an equation, which is Bayes' Theorem. It lets you take the test results and correct for the "skew" introduced by false positives. You get the real chance of having the event. Here's the equation:

$$\mathrm{Pr}(\mathrm{A}|\mathrm{X}) = \frac{\mathrm{Pr}(\mathrm{X}|\mathrm{A})\,\mathrm{Pr}(\mathrm{A})}{\mathrm{Pr}(\mathrm{X}|\mathrm{A})\,\mathrm{Pr}(A) + \mathrm{Pr}(\mathrm{X}|\sim\mathrm{A})\,\mathrm{Pr}(\sim A)}$$

And here's the decoder key to read it:

- Pr(A|X) = Chance of having cancer (A) given a positive test (X). This is what we want to know: How likely is it to have cancer with a positive result? In our case it was 7.8%.
- Pr(X|A) = Chance of a positive test (X) given that you had cancer (A). This is the chance of a true positive, 80% in our case.
- Pr(A) = Chance of having cancer (1%).
- Pr(~A) = Chance of not having cancer (99%).
- Pr(X|~A) = Chance of a positive test (X) given that you didn't have cancer (~A). This is a false positive, 9.6% in our case.

Try it with any number:

Bayes Theorem

R1	Actual_probability = 1	0.01
R2	Prob_true_positive = 8	0.8
R3	Prob_false_positive =	0.096
R4		Chance positive test means positive result
R5		7.76397515528 %

+5 rowsClear
instacalc

It all comes down to the chance of a **true positive result** divided by the **chance of any positive result**. We can simplify the equation to:

$$\Pr(\mathrm{A}|\mathrm{X}) = \frac{\Pr(\mathrm{X}|\mathrm{A})\Pr(\mathrm{A})}{\Pr(\mathrm{X})}$$

Pr(X) is a normalizing constant and helps scale our equation. Without it, we might think that a positive test result gives us an 80% chance of having cancer.

Pr(X) tells us the chance of getting *any* positive result, whether it's a real positive in the cancer population (1%) or a false positive in the non-cancer population (99%). It's a bit like a weighted average, and helps us compare against the overall chance of a positive result.

In our case, Pr(X) gets really large because of the potential for false positives. Thank you, normalizing constant, for setting us straight! This is the part many of us may neglect, which makes the result of 7.8% counter-intuitive.

Intuitive Understanding: Shine The Light

The article mentions an intuitive understanding about shining a light through your real population and getting a test population. The analogy makes sense, but it takes a few thousand words to get there :).

Consider a real population. You do some tests which "shines light" through that real population and creates some test results. If the light is completely accurate, the test probabilities and real probabilities match up.

Everyone who tests positive is actually "positive". Everyone who tests negative is actually "negative".

But this is the real world. Tests go wrong. Sometimes the people who have cancer don't show up in the tests, and the other way around.

Bayes' Theorem lets us look at the skewed test results and correct for errors, recreating the original population and finding the real chance of a true positive result.

Bayesian Spam Filtering

One clever application of Bayes' Theorem is in spam filtering. We have

- Event A: The message is spam.
- Test X: The message contains certain words (X)

Plugged into a more readable formula (from Wikipedia):

$$\Pr(\text{spam}|\text{words}) = \frac{\Pr(\text{words}|\text{spam})\,\Pr(\text{spam})}{\Pr(\text{words})}$$

Bayesian filtering allows us to predict the chance a message is really spam given the "test results" (the presence of certain words). Clearly, words like "viagra" have a higher chance of appearing in spam messages than in normal ones.

Spam filtering based on a blacklist is flawed — it's too restrictive and false positives are too great. But Bayesian filtering gives us a middle ground — we use *probabilities*. As we analyze the words in a message, we can compute the chance it is spam (rather than making a yes/no decision). If a message has a 99.9% chance of being spam, it probably is. As the filter gets trained with more and more messages, it updates the probabilities that certain words lead to spam messages. Advanced Bayesian filters can examine multiple words in a row, as another data point.

Further Reading

There's a lot being said about Bayes:

- Bayes' Theorem on Wikipedia
- Discussion on coding horror
- The big essay on Bayes' Theorem

Have fun!

Other Posts In This Series

1. A Brief Introduction to Probability & Statistics

May 6, 2007 Math Printable version

88 Comments

1. Gavrilo Princep May 7, 2007 at 3:40 am

you have a typo, in …

9.6% of mammograms miss breast cancer when it is there (and therefore 90.4% say it is there when it isn't).

… you meant to say somthing like :

9.6% of mammograms incorrectly indicate breast cancer when it isn't there, and the other 90.4% correctly say it is not there when, well, it is not there.

2. Kalid May 7, 2007 at 10:42 am

Thanks Gavrilo — I just fixed it.

3. Amal May 12, 2007 at 11:48 am

Hey, here's an interesting bayes problem i came across first in a book (The curious incident of the dog in the night time).

Suppose you are in a game show. You are given the choice of three doors – one of which conceals a valuable prize and the
others conceal a goat.

After you make a choice, the host opens one of the other doors (–one without a prize).

He then gives you the option of staying with the initial choice of door or switching to the other door. The door finally chosen is then opened.

Should you switch, not switch, or does it make no difference what the contestant does?

ANSWER
by Bayes theorum you can see that if you switch u'd have a 2:1 advantage.

4. Kalid May 13, 2007 at 6:27 pm

Hi Amal, thanks for dropping by. Yes, I like that question too, it was presented to us as "The Monty Hall" problem when studying computer science.

It's pretty amazing how counter-intuitive the results can be — switching your choice after you've picked "shouldn't" change your chances, right? I plan on writing about this paradox, too 🙂

5. Ed August 27, 2007 at 10:49 pm

Oddly useful! I've been reading Bayes explanations for a while, and this one really hit home for me for some reason.

One thing that you might consider adding (something I've never seen) is a pie-chart visualization of what's going on. Basically, you have a pie of 100% of people. 1% of that pie has cancer, so that's a tiny slice. The test will produce a positive for 80% of that 1% slice + 9.6% of the remaining 99% slice– you can imagine that as a little blue translucent piece of appropriate size that covers most of the 1% slice and a chunk of the 99% slice. From that mental image, it's obvious what's going on– there's a lot more blue on the 99% than on the 1%. Might be too complicated, but hey. 🙂 Anyways, thanks.

6. Kalid August 30, 2007 at 1:29 am

Hey Ed, thanks for the comment. I agree — some type of chart may make the relationship that much clearer. Appreciate the suggestion, I'll put one together.

7. Lee October 21, 2007 at 3:50 pm

Bayes theorem can also be thought of as

True Positives
————————————–
True Positivees + False Positives

So a large number of false positives reduces the accuracy of the test because the denominator increases.

8. Kalid October 21, 2007 at 9:23 pm

Thanks Lee! That's a great way to put it.

9. Randy November 7, 2007 at 12:21 am

About Monty Hall- the Bayes application to this seems very forced. The Monty Hall problem is a simple probability problem, or it can be viewed as a partitioning problem. See: http://randy.strausses.net/tech/montyhall.htm

Using Bayes for this makes it needlessly complex, not "betterExplained".

Similarly, the article above is needlessly complex- nuke the first equation and leave the simpler one. You just pulled it out of thin air anyway- it doesn't help anyone.

The usual diagram, given in HS stats classes, is a rectangle, with A, ~A on the top, B, ~B on the side. Say A is .9 and B is .2. The area of the small quadrant (.02), is the probability of A and B both happening. This area can be also viewed as P(A|B)*P(B) or P(B|A)*P(A). You have to explain why, but it's pretty evident from the diagram. Then just equate these two and divide by P(B) and you have the simpler equation.

10. Kalid November 8, 2007 at 2:29 pm

Hi Randy, Bayes may be overkill for the Monty Hall problem, but it's interesting to see that it can apply there as well.

Yes, the diagram you mention may be a helpful addition to the discussion above, appreciate the feedback.

11. numerodix November 27, 2007 at 1:20 pm

Just wanna say thank you for writing this. I know about the original article and I tried reading it but somewhere along the way I got lost and couldn't follow it.

12. Kalid November 27, 2007 at 1:51 pm

Hi numerodix, you're welcome — I found the original article interesting but a bit long as well, so I decided to summarize it here.

13. Matteo April 7, 2008 at 2:28 am

Hello, I just came upon this site and I'm finding it beautiful. I think I spotted an error in this article, though.
When you say:

"”Of those 100, 80 will test positive” rather than “80% of the 1% will test positive”).",

you probably wanted to say: “rather than ’80% of the 100% will test positive’”.

14. Kalid April 8, 2008 at 12:40 am

Hi Matteo, thanks for the comment. The statement actually refers to the original 1%, so it’s giving a way of giving compound percentages (80% of 1% vs. 80 out of 10,000).

15. Anonymous May 6, 2008 at 5:47 am

wow thank you so much for this, you really did a good job explaining it, i have my AP statistics exam today at noon so this might save me 😀

16. John D Stackpole January 11, 2009 at 8:26 pm

Randy, back on Nov 7 2007, suggested using overlapping rectangles – Venn diagrams – to help clarify the Rev. Bayes. In their book “Chances Are…” (Viking Penguin, 2006), Kaplan & Kaplan did so on pp. 184 ff. Indeed it does help.

17. Anonymous April 15, 2009 at 11:32 pm

you. are. the. best.

18. patty April 17, 2009 at 7:50 am

oh my god this is the dogs bollocks for my molecular phylogenetics revision!

19. Kalid May 5, 2009 at 3:19 pm

@Anonymous: Thanks!

@John: Appreciate the reference. Another explanation with a venn diagram: http://blog.oscarbonilla.com/2009/05/visualizing-bayes-theorem/

@Anonymous: Thank you!

@Patty: Glad it helps 😀

20. Dan Weisberg October 2, 2009 at 10:40 am

This is one of the best explanations I've found. Perhaps we can see if I really understand it by trying a real world problem I'm wrestling with.

Here's the data:
– The odds of a chest pain (CP) being caused by a heart attack is 40%.
– The odds of a CP being caused by other factors (anxiety, depression, etc.) is 60%.
– The odds of a heart attack occurring to a female above age 50 is 80%.
– The odds of a heart attack occurring to a female under age 50 is 20%.

I am presented with a 24 year old female who says she is having chest pain. What is the probability that her chest pain is caused by a heart attack? Is it 0.4 x 0.2 = 0.08?

Also, 78% of patients having heart attacks present with diaphoresis (sweating), so 22% of patients having heart attacks don't sweat. This female is not sweating, so are the odds of her having a heart attack 0.22 x 0.08 = 0.0176?

Thank you!

21. Emily Riley October 29, 2009 at 2:58 am

Thanks for writing this!! Even my stats prof was making this too difficult for everyone, but you have simplified it for me. I now have an understanding of the Bayes formula (enough to write my midterm this morning 😃).

22. AYUSH December 19, 2009 at 2:45 am

thanks! finally got the concept behind bayes rule

23. Kalid December 21, 2009 at 11:48 pm

@Ayush: Glad it helped!

24. Andrew January 7, 2010 at 10:49 am

Could you work out an example of an email with *two* words, say 'Viagra' and 'hello'?

25. Aditya March 25, 2010 at 1:13 am

I didnt get that, bayes theorem is still a tilted pot for me, but thanks for helping!

26. Asim May 7, 2010 at 8:18 am

what would happen if we have to consider other prior probability..lets say, the doctor looked at the symptoms of the patient and guessed that he has 60 percent chance of having cancer. Doctor sends him for the test and test showed positive result. How would we incorporate that 60 percent odd of having cancer based on the patient's symptoms to the Bayesian equation.

27. Francis August 29, 2010 at 12:34 pm

hi
thank you for this article. The first time I came across Bayes Theorem in a business statistics book it was not so clear at all. No it makes more sense for me and its pretty clear..

28. Kalid August 29, 2010 at 11:22 pm

@France: You're welcome, glad it helped. I understand it better now too, but there's still more to go before it's completely intuitive for me :). I'd like to do a follow-up to this focusing on using the probabilities it predicts.

29. kr December 7, 2010 at 9:46 am

On the cancer example, it's interesting to see that a negative test is really significant. That is, if the test says you don't have cancer, then probability of not having cancer is 99.78% ! So, the value of mammogram is that the healthcare $ can be employed in further investigating the positive (+ and -) cases only.

30. HS January 2, 2011 at 10:40 am

Thank you for taking the time to write this – it has really helped me get my head around the concept!

31. Anonymous May 1, 2011 at 7:28 pm

@Dan Weisberg: 14% of chest pains in women under 50 are therefore caused by heart attacks. Its (0.2*0.4)/((0.2*04)+(0.8*0.6)) = 0.14

32. Anonymous May 2, 2011 at 12:52 am

@Dan Weisberg: for some odd reason, my posts seem to disappear and reappear… anywho, the other part to your question (as earlier posted) is 4.5% irrespective of when you choose to include the diaphoresis variable.

33. Bips May 29, 2011 at 3:38 pm

wow, thank you so much for this, made the concept so much clearer to me

34. Kalid May 29, 2011 at 9:45 pm

@Bips: Glad it helped!

35. Concerned. June 29, 2011 at 1:54 pm

This post is filled with jargon, and I'm surprised people can tolerate it.

36. Pooja August 7, 2011 at 1:36 pm

Nicely explained; this really helped me study for my stat final. Thanks!

37. Alec August 10, 2011 at 5:55 pm

Thanks so much for this simple breakdown!

38. Kalid August 10, 2011 at 7:34 pm

@Concerned: You should have seen the original :).

@Pooja, @Alec: Thanks!

39. One Puzzle Piece Short September 10, 2011 at 8:52 pm

"A Bayesian is someone who, vaguely expecting a horse, and glimpsing the tail of a donkey, concludes he has probably seen a mule." – John Hussman

Not really helpful, but rather funny.

40. MH November 11, 2011 at 10:06 pm

Extremely helpful article. The lecture by the stats prof was utterly bewildering….. thanks!

41. kalid November 11, 2011 at 10:44 pm

@MH: Awesome, glad it helped.

42. Joseph Senecal November 15, 2011 at 12:56 pm

Your explanation using the cancer example was superb.

43. kalid November 15, 2011 at 1:33 pm

@Joseph: Thanks!

44. Bruce Bohm November 23, 2011 at 3:28 pm

Do you know of "The theory that would not die," Sharon Bertsch McGrayne's history of the Bayes controversy? It is history at its best including breaking the Enigma Code during the Second World War, plus lots of other applications. The book was published by Yale University Press in 2011.

45. Zachary Sexton December 30, 2011 at 2:12 pm

Woman's percent chance of having cancer if tested negative for breast cancer.

P(False Neg)/(P(False Neg) + P(Accurate Negative)) => .01(.2)/(.01(.2) + .99(.904)) =.0022 = .22%

How's my math?

46. Jonathan December 31, 2011 at 9:40 am

Very Helpful! I appreciate the intuitive explaination. Thank You!

47. Richard Karpinski January 3, 2012 at 12:02 am

Your analysis of the Monty Hall problem *assumes* that the host's intentions and attitudes are irrelevant to his *choice* of whether to reveal the secret of one non-winning selection. If the host wishes you to win, he will only open a door when you chose the wrong one first, but if he wishes you to lose, then he will open a door only when you have chosen the right one first. Thus unless you know that his wishes don't matter then Bayes' Theorem will not help. If you were to know his wishes do matter, then your rational choice depends solely on your estimate of whether he wishes you to win (switch) or lose (don't switch).

48. kalid January 3, 2012 at 6:23 pm

@Richard: It might need to be clarified, but the rules are the host will always reveal a goat in one of the two doors you didn't pick. Even if you've already picked the car, he will still reveal a goat.

49. Geoff February 8, 2012 at 8:03 pm

Thanks for writing that! Really cleared it up…I have also since read the explanation with a small binomial tree, which was pretty helpful, but this really hit home. Thanks for the clear explanation!

50. kalid February 9, 2012 at 10:50 pm

@Geoff: Glad it helped! I'm planning on doing a follow-up to Bayes since it's used so much in spam filtering and other machine learning.

51. MANWARI G. ALEX March 14, 2012 at 10:46 am

bayes theorem is incredible

52. Günther March 26, 2012 at 1:10 pm

Hello Kalid, I am glad I found this helpful site and also your older one at princeton.edu. About your explanation of Bayes' theorem, however, I have two minor caveats.

You wrote: "Interesting — a positive mammogram only means you have a 7.8% chance of cancer, rather than 80% (the supposed accuracy of the test)."

Shouldn't we better take 90.4% as the supposed accuracy of the test? The sensitivity of the test is 80%; it tells us that 20% of the negative results are false; but if we've got a positive mammogram, that information about negative results doesn't interest us in the first place. Much more interesting is the specificity of the test: it's 90,4% and therefore tells us that 9,6% of the positive results will be false. If our test result is positive, we would wish the specificity of the test were still much lower, because then we could have a much better founded hope of being healthy despite the positive result.

You wrote: "It all comes down to the chance of a true positive result divided by the chance of any positive result. […] Pr(X) is a normalizing constant and helps scale our equation. Without it, we might think that a positive test result gives us an 80% chance of having cancer. Pr(X) tells us the chance of getting any positive result, whether it's a real positive in the cancer population (1%) or a false positive in the non-cancer population (99%). It's a bit like a weighted average, and helps us compare against the overall chance of a positive result. In our case, Pr(X) gets really large because of the potential for false positives. Thank you, normalizing constant, for setting us straight!"

Instead of 80% I would again prefer to read 90,4% here. But even after this modification I doubt whether your explanation of the "normalizing" could be called sufficient, because in my interpretation it's not only our using Pr(X) as denominator that helps us "scaling the equation". We've got a positive test result, and that's why we are inquiring after Pr(A|X). As soon as we take Pr(X) into consideration, we must not compare it with the specificity of the test nor with the unqualified (i.e., unconditional) sensitivity; rather, we have to compare it with the conditional probability of the following case: that someone has breast cancer (1%) and that the test detects it (80%). So not only the denominator Pr(X), but also the factor Pr(A) in the numerator (i.e., the a priori probability of having breast cancer) is a "normalizing" element, isn't it?

53. gerry worts March 28, 2012 at 8:23 pm

Here are two contradictory statements in your article regarding False positives. I quote:

Pr(X|~A) = Chance of a positive test (X) given that you didn't have cancer (~A). This is a false positive, 9.6% in our case.

The chances of a false positive = chance you don't have cancer * chance test caught it anyway = 99% * 9.6% = 0.09504

These two interpretations of false positives are often mixed in various texts.
One is Pr(X|~A) and the other is Pr(X∩~A)

54. elmira komijani May 12, 2012 at 5:45 am

Hello guys;
could you help me to solve this problem as soon as possible?
Transplant operations for hearts have the risk that the body may reject the organ. A new test has been developed to detect early warning signs that the body may be rejecting the heart. However, the test is not perfect. When the test is conducted on someone whose heart will be rejected, approximately two out of ten tests will be negative (the test is wrong). When the test is conducted on a person whose heart will not be rejected, 10% will show a positive test result (another incorrect result). Doctors know that in about 50% of heart transplants the body tries to reject the organ.
*Suppose the test was performed on my mother and the test is positive (indicating early warning signs of rejection). What is the probability that the

body is attempting to reject the heart?
*Suppose the test was performed on my mother and the test is negative (indicating no signs of rejection). What is the probability that the body is attempting to reject the heart?

55. DENNIS June 28, 2012 at 5:16 am

MMMMHHH… BEEN HELPFULL

56. CollegeStudent July 10, 2012 at 7:21 am

Wow, thank you so much for this. I looked at several examples and tutorials before arriving here and your use of tables really, really helped wrap my head around it. Thanks!!

57. Carlos August 28, 2012 at 9:23 pm

Excellent!

58. Anonymous August 28, 2012 at 10:42 pm

Hey guys great work. Thank you

59. Anonymous November 13, 2012 at 11:24 pm

You might want to look at this if you want to understand Bayes' theorem in less than 2.5 minutes:
http://www.youtube.com/watch?v=D8VZqxcu0I0

60. oracleaide December 28, 2012 at 11:16 am

I think a Venn Pie chart (with overlapping sectors) could really help to make an intuitive explanation:

http://oracleaide.wordpress.com/2012/12/26/a-venn-pie/

61. Anonymous January 6, 2013 at 5:43 pm

a graphic solver for bayes' theorem… again – using venn pie chart:
http://dl.dropbox.com/u/133074120/venn_pie_solver.html

62. Ada March 13, 2013 at 4:39 pm

I can not even explain the feelings of gratitude and affection I now have towards you. Thank you so much.

63. kalid March 13, 2013 at 5:01 pm

Hi Ada, really glad it helped!

64. thekfactor March 22, 2013 at 1:33 pm

This is a wonderful site, and I try and visit your site whenever possible…Excellent work you are doing here! Keep up the awesome work!

65. Steve Bithell March 30, 2013 at 1:21 pm

This is just so bloody good!
What a fantastic bit of insight, and I used it to explain the nature of testing to a room full of people the other day.
Sometime, you'll stop writing these articles and I (and many, many others) will be stuffed.
Keep it up Kalid

66. kalid April 3, 2013 at 3:05 pm

Hi Steve, glad it's clicking! Love that you were able to help others learn. Appreciate the encouragement, hope to keep going as long as I can ;).

67. Nat Napoletano April 4, 2013 at 7:51 am

I agree with the way that this article was presented. Sometimes people want to see where the subject is going before they invest the time in understanding the math. I have put together a fun series of videos on YouTube entitled "Bayes' Theorem for Everyone". They are non-mathematical and easy to understand. And I explain why Bayes' Theorem is important in almost every field. Bayes' sets the limit for how much we can learn from observations, and how confident we should be about our opinions. And it points out the quickest way to the answers, identifying irrationality and sloppy thinking along the way. All this with a mathematical precision and foundation. Please check out the first one:

http://www.youtube.com/watch?v=XR1zovKxilw

68. Dave April 26, 2013 at 4:13 am

Hi. Nicely condensed article. However, unless I'm interpreting your conclusion incorrectly, you stated that 'a positive test will result in only 1/11 people having cancer (7.8% to be exact)'. My math tells me 1/11 is 9.1%! Since we know 7.8% is correct then the probability is actually 1/13. If we normalise 80 true positives out of 10,000 to be 1 true positive out of 125 then out of the 1030/8 = 12.875 (or 13, not 11) with positive mammograms, 1 will have cancer.

69. Kalid April 26, 2013 at 9:01 am

Hi Dave, thanks for the comment. I should have clarified that part — it should be more of an "If you eyeball the probability table (without a calculator), what's your conclusion?" (My goal for eyeballing is to test our intuition without explicitly computing the formula, just like we might "eyeball" the square root of 20 as 4.5.)

Eyeballing the table, if we have 100 people, about 1 will have cancer and be detected (technically "0.8″), and about 10 will not have cancer and be detected (99 people * 9.6% of false positive). So we have roughly 1 real cancer event, and 10 false ones, for an estimate of 1/11 chance of the event being meaningful. (The actual amount is 7.8% as you said, because the numbers aren't 1 and 10 exactly).

Really appreciate the comment, that part wasn't clear and I'll clean it up.

70. John June 7, 2013 at 12:01 pm

I like the way you think about math, and your approach — really useful!

Can you play teacher correcting my work for a moment? I took your approach and applied it to 2 other Bayesian examples from other sites — the NYT article that the "big" essay references, and another from techtarget:
http://whatis.techtarget.com/definition/Bayesian-logic

The application against the NYT example worked fine (the possibility of a an unfair coin, given the 3 coin tosses). I assume the author slightly rounded the result for simplicity — revised posterior probability goes from 1 in 3 to 4 in 5 (or 80%), whereas I got 79.758%).

However, in the techtarget example, I don't see how the author gets the 50% probability. Here's the salient part of the problem restated:

———————————-
"[S]uppose that we have a covered basket that contains three balls, each of which may be green or red. In a blind test, we reach in and pull out a red ball. We return the ball to the basket and try again, again pulling out a red ball. Once more, we return the ball to the basket and pull a ball out – red again. We form a hypothesis that all the balls are all, in fact, red. Bayes' Theorem can be used to calculate the probability (p) that all the balls are red (an event labeled as "A") given (symbolized as "|") that all the selections have been red (an event labeled as "B"):

p(A|B) = p{A + B}/p{B}

Of all the possible combinations (RRR, RRG, RGG, GGG), the chance that all the balls are red is 1/4; in 1/8 of all possible outcomes, all the balls are red AND all the selections are red. Bayes' Theorem calculates the probability that all the balls in the basket are red, given that all the selections have been red as .5…"
——————————-

I calculate as follows:

p(A) = 0.25 — i.e., RRR, all red balls in hat
p(~A) = 0.75 — chance hat contains RRG, RGG or GGG
p(B|A) = 1.0 — chance that the 3 selections are red, given the hat contains RRR
p(B|~A) = 0.125 — chance of picking 3 reds, given the hat contains RRG, RGG, or GGG.
p(B|~A) derived as:

– Pick R on first ball = 0.5 (R or G); pick RR after 2 balls = 0.25 (RR, RG, GR or GG); pick RRR after 3 balls = 0.125 (RRR, RRG, RGG, GGG, GRR, GGR, RGR, GRG).
– RRG: 0.67 * 0.125 = 0.08375
– RGG: 0.33 * 0.125 = 0.04125
– GGG: 0
– RRG + RGG + GGG = 0.08375 + 0.04125 + 0 = 0.125

Execute Bayes:
– Numerator: p(B|A) * p(A) = 1 * 0.25 = 0.25
– Denominator: p(B|A) * p(A) + p(B|~A) * p(~A) = (1 * 0.25) + (0.125 * 0.75) = 0.25 + 0.09375 = 0.34375
– Result: 0.25 / 0.34375 = 0.72727

The author gets 0.5. Using an abbreviated form of Bayes in her text, she has it that p(A + B) = 0.125 — "in 1/8 of all possible outcomes, all the balls are red AND all the selections are red. " I struggle with that statement, but going with it, I take that as meaning all TRUE positives. And, p(B) = the event where all 3 selections are red, whether the hypothesis of the hat containing RRR is true or not. If her dividend is 0.125, in order to obtain 0.5, she would need a divisor of 0.25 (0.125 / 0.25 = 0.5).

This would imply p(B|~A) * p(~A) = 0, which is not possible (i.e., the chance of picking 3 red balls where the possible population does indeed contain the possibility of red balls existing in some proportion — RRG or RGG).

Help!

Thank you! (Really — Thank you!!)

71. anonymus June 9, 2013 at 12:28 am

Very well written . Very helpful

72. kalid June 13, 2013 at 1:50 pm

@John: Interesting question. Before going further though, I think the techtarget author made some mistakes. If the balls are randomly distributed (between Red and Green) then the 8 possible outcomes are

RRR RRG RGR RGG GRR GRG GGR GGG

Sure, RRG may look the same as RGR and GRR, but that combo (2 reds and 1 green) has a 3/8 chance of happening. Similar for GGR. So the real odds are:

All reds = 1/8
2 reds, 1 green = 3/8
2 greens, 1 red = 3/8
All greens = 1/8

From there, I'd probably have to work out a quick table to see the probabilities. The idea is to figure out how many false positives you get, how many true positives you get, and find the chance of the event as (true positives) / (false positives + true positives).

@Anon: Thanks.

73. Exori August 18, 2013 at 3:06 pm

Thank you, makes sense now!
The data you used with the mammogram example, were the probabilities in it true for tests actually in use?

74. Martin Ferris August 23, 2013 at 2:28 pm

That is an awesome explanation. thankS!

75. Dave November 24, 2013 at 7:38 pm

You have a mistake there. 1% * 80% is not = 0.08, BUT it is equal to 0.008

Check it out here on Wolfram Alpha: http://www.wolframalpha.com/input/?i=1%25+*+80%25

76. Kalid November 25, 2013 at 2:33 am

Hi Dave, I searched but don't see a reference to "0.08″ in the article — more than happy to correct if there's a calculation error.

77. Savitri KVL January 28, 2014 at 1:12 am

Hi,

In Youtube, I found a good lecture on this theorem.

http://www.youtube.com/user/InsofeVideos/videos?view=46&shelf_id=9&tag_id=UCJ1R2JZ_ecOBxyCiURdmgAQ.3.cpee&sort=dd

Check it out.

78. Ron Parkinson March 7, 2014 at 1:24 pm

Your explanation and example was really helpful!!! You should urge your readers to proactively use the calculator you have provided to explore what happens as "false alarms" go up and "missed detections" go down. One can think of kinds of tests where some threshold number can be varied to tailor the Pr(X|A) 's, say, if you don't mind false alarms as a cost of being sure that missed detections are very low…

79. kalid March 9, 2014 at 12:06 am

Thanks Ron, glad you liked it! Great point, the idea behind Bayes is to account for the false alarm / missed detection tradeoff that a test will have. (False alarm / missed detection is a great way to explain false positive and false negative also.)

80. Todd March 18, 2014 at 7:55 am

Kalid,
Suppose that someone approaches you with a friend who she says has a photographic memory, i.e., eidetic imagery (EI). The reason she comes to you is that you have an infallible test of EI, but it has a .05 rate of false positives: that is, if a person has EI, the test will come up positive 100% of the time, but if the person does not have EI, the test will come up positive 5% of the time. Suppose also that only one person in 10,000 has EI. So you test this person's friend, and the test is positive for EI. What is the probability that the friend actually has EI?

81. kalid March 18, 2014 at 7:58 am

Hi Todd, for this one, I'm going to let you work it out :).

Suppose you have 10,000 people in an auditorium.

1) How many actually have EI? (call this R, for real)
2) How many test positive for having EI, but do not have this? (call this F, for fake)

The probability of actually having it would be R / (R + F) [the number of real people in the total population who tested positive]

82. Chris Gundersen March 19, 2014 at 10:49 am

Superb article, very well explained. It's been a couple of years since I last studies Bayes' theorem, and this was a very useful refresher.

83. Todd March 21, 2014 at 7:03 am

so .001 for R
and .05 for F

so .001/.001+.05= .0196
is this right?

84. Doc John March 23, 2014 at 2:56 pm

Excellent article Kalid, reminded me succinctly what Bayes is all about. Are you being hoodwinked into doing people's homework though!
Great work.

85. Joesph Mc Glade March 25, 2014 at 2:00 am

Kalid you're fantastic, I found this site while reading how they recovered the Air France Plane. I wonder Kalid do you have a bet on sports,I'm 70yrs old this October and like thosands of others we try to get our day in by going into bookmakers shops.You're insight would help fellow punters as the bookmakers have it all there own way,they have 18 screens showing racing from greyhound tracks,horse racing from around the world also now virtual horse racing,virtual horse racing,football from around the world and worse of all they have gaming machines.Just to end from the 18 screens there is only one that gives you the results.Thanks for your insight.

86. Kamille March 28, 2014 at 1:52 pm

Kalid, thank you for these well-written article. I am presenting a poster on an medical algorithm that calculated ratios based on Bayes' Theorem. If I am asked how the ratios were calculated, I'll be more prepared. Thanks again!

87. kalid March 28, 2014 at 1:59 pm

Thanks Kamille, glad it helped!

88. kalid March 28, 2014 at 2:02 pm

Thanks Joseph, I appreciate it! Ah, I don't know much about sports betting, but I do know that odds are a more natural way to express probabilities than a percentage :).